



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/818,607	03/28/2001	Yasuo Okutani	35.G2761	1901

5514 7590 12/03/2003

FITZPATRICK CELLA HARPER & SCINTO
30 ROCKEFELLER PLAZA
NEW YORK, NY 10112

EXAMINER

WOZNIAK, JAMES S

ART UNIT	PAPER NUMBER
----------	--------------

2655

DATE MAILED: 12/03/2003

¢

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary

Application No.

09/818,607

Applicant(s)

OKUTANI ET AL.

Examiner

James S. Wozniak

Art Unit

2655

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 03/28/2001.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-25 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-25 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 28 March 2001 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
- Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. §§ 119 and 120

- 12) ☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☒ All b) ☐ Some * c) ☐ None of:
1. ☒ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- * See the attached detailed Office action for a list of the certified copies not received.
- 13) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application) since a specific reference was included in the first sentence of the specification or in an Application Data Sheet. 37 CFR 1.78.
- a) ☐ The translation of the foreign language provisional application has been received.
- 14) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121 since a specific reference was included in the first sentence of the specification or in an Application Data Sheet. 37 CFR 1.78.

Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☒ Information Disclosure Statement(s) (PTO-1449) Paper No(s) 2.
- 4) ☐ Interview Summary (PTO-413) Paper No(s). _____.
- 5) ☐ Notice of Informal Patent Application (PTO-152)
- 6) ☐ Other: _____.

Detailed Action

Double Patenting

1. The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. See *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); *In re Van Ornum*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); and, *In re Thorington*, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) may be used to overcome an actual or provisional rejection based on a nonstatutory double patenting ground provided the conflicting application or patent is shown to be commonly owned with this application. See 37 CFR 1.130(b).

Effective January 1, 1994, a registered attorney or agent of record may sign a terminal disclaimer. A terminal disclaimer signed by the assignee must fully comply with 37 CFR 3.73(b).

2. Claims 1-24 are provisionally rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1-21 of copending Application No. 09/818,581. Although the conflicting claims are not identical, they are obvious variations of one another because both disclose a speech synthesis system utilizing concatenation and modification distortion in the selection a best instance of a speech unit, and Application No.

09/818,581 employs the use of Nbest processing in the selection process, a difference which, to one of ordinary skill in the art, at the time of invention, would have been obvious since an Nbest processing method would be more time effective in selecting a best instance. Nbest processing allows only the best speech candidates to be examined instead of an entire speech database, thus resulting in a reduction in processing time.

This is a provisional obviousness-type double patenting rejection because the conflicting claims have not in fact been patented.

Claim Rejections - 35 USC § 102

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(a) the invention was known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the invention thereof by the applicant for a patent.

4. Claims **1, 2, 4-6, 12-14, 16-18, and 24** are rejected under 35 U.S.C. 102(a) as being anticipated by U.S. Patent: 5,913,193 to Huang et al.

With respect to **Claims 1 and 13**, Huang discloses:

A speech signal processing apparatus and method for performing speech synthesis by concatenating a plurality of selected synthesis units and modifying the synthesis units based on predetermined prosody parameters (*Fig. 5, concatenation of speech units based upon prosodic parameters of an input text and speech waveform synthesis performed thereafter, Col. 9, Lines 49-56*), said apparatus comprising:

Distortion obtaining means for obtaining a distortion, which may be generated from selection to synthesis of the synthesis units (*determining distortion in the process of speech unit selection as seen in Fig. 7, Element 178*);

Selection means for selecting synthesis units to be used for speech synthesis, based on the distortion obtained by said distortion obtaining means (*best instance speech unit selection based upon distortion, Col. 9, Lines 44-43, Fig. 7, Element 182*); and

Speech synthesis means for performing speech synthesis based on the synthesis units selected by said selection means (*speech synthesis waveform generation upon selection of a best speech unit, Col. 9, Lines 49-53, and speech synthesizer, Fig. 1, Element 36*).

With respect to **Claims 2 and 14**, Huang recites:

An apparatus according to claims 1 and 13, respectively, wherein said selection means selects a plurality of synthesis units based on a phoneme series including a plurality of phonemes (*selection of a best instance of each speech unit composed of a phonetic string, Col. 4, Lines 48-53*).

With respect to **Claims 4 and 16**, Huang recites:

An apparatus according to claims 1 and 13, respectively, wherein said selection means selects the synthesis units to be used in speech synthesis so as to minimize the distortion (*Col. 2, Lines 1-7*).

With respect to **Claims 5 and 17**, Huang discloses:

An apparatus according to claims 1 and 13, respectively, wherein said distortion obtaining means obtains the distortion based on a concatenation distortion generated by concatenating a synthesis unit to another synthesis unit and a modification distortion generated

Art Unit: 2655

by modifying the synthesis unit (*spectral distortion between adjacent instances and prosodic distortion, Col. 3, Lines 1-6*).

With respect to **Claims 6 and 18**, Huang suggests:

An apparatus according to claims 1 and 13, respectively, wherein said distortion obtaining means uses a value obtained by adding (*summing the distortions of an instance sequence, Col. 9, Lines 44-47*) a concatenation distortion generated by concatenating a synthesis unit to another synthesis unit and a modification distortion generated by modifying the synthesis unit as the distortion (*spectral distortion between adjacent instances and prosodic distortion, Col. 3, Lines 1-6*).

With respect to **Claims 12 and 24**, Huang discloses:

An apparatus according to claims 1 and 13, respectively, further comprising:

Input means and step for inputting text data (*computer terminal for data input, Col. 3, Lines 14-18*);

Language analysis means and step for performing language analysis of the text data (*natural language processor, Col. 4, Lines 24-26, Fig. 1, Element 32*); and

Prosody-parameter generation means and step for generating the predetermined prosody parameters based on a result of analysis of said language analysis means and step (*prosody engine, Col. 4, Lines 39-44, Fig. 1, Element 35*).

Thus, Huang anticipates the invention as recited in Claims 1, 2, 4-6, 12-14, 16-18, and 24.

Claim Rejections - 35 USC § 103

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

6. **Claims 3, 7-9, 15, and 19-21** are rejected under 35 U.S.C. 103(a) as being unpatentable over Huang in view of U.S. Patent: 6,366,883 to Campbell.

With respect to **Claims 3 and 15**:

Huang teaches the speech synthesis apparatus and method that utilizes an obtained distortion, corresponding to a speech unit, in selecting a best instance in speech concatenation as applied to Claim 1, but does not teach: synthesis units corresponding to a single phoneme that are selected based upon the distortion associated with that particular frame as recited in Claim 3.

With respect to **Claims 3 and 15**, Campbell discloses:

An apparatus according to claims 1 and 13, respectively, wherein said distortion obtaining means obtains a distortion which may be generated in each of a plurality of synthesis units corresponding to one phoneme, and wherein said selection means selects one synthesis unit from among the plurality of synthesis units corresponding to the one phoneme (*speech unit selection process of an N best group of individual phonemes, Fig. 7*).

Huang and Campbell are analogous art because they are from a similar field of endeavor in speech synthesis and concatenation. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the partitioning of individual

speech units as individual phonemes in a best unit selection process as taught by Campbell with the best instance selection method utilizing distortion as taught by Huang to create a speech synthesis method capable of producing a high quality speech sequence with minimized distortion on an individual phoneme basis. Therefore, it would have been obvious to combine Campbell with Huang for the benefit of obtaining a speech synthesis system capable of producing high quality synthesized speech by selecting individual phonemes for concatenation based upon distortion, to obtain the invention as specified in Claims 3 and 15.

With respect to **Claims 7 and 19**:

Huang teaches the speech synthesis apparatus and method that utilizes an obtained distortion, corresponding to a speech unit, in selecting a best instance in speech concatenation as applied to Claims 1 and 13 and also a summing of distortions in the speech unit selection process (*Col. 9, Lines 44-48*), but does not teach a weighted sum of distortion as recited in Claim 7.

With respect to **Claims 7 and 19**, Campbell discloses:

An apparatus according to claims 3 and 17, respectively, wherein said distortion obtaining means calculates the distortion as a weighted sum of the concatenation distortion and the modification distortion (*selecting a speech unit based upon a weighted coefficient vectors, Col. 2, Lines 37-38*).

Huang and Campbell are analogous art because they are from a similar field of endeavor in speech synthesis and concatenation. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the use of a weighted coefficient in the process of best speech unit selection as taught by Campbell with the best instance selection method utilizing a distortion sum as taught by Huang to provide a means of minimizing

Art Unit: 2655

concatenation cost expressed through a weighing function and thus providing higher quality synthesized speech. Therefore, it would have been obvious to combine Campbell with Huang for the benefit of obtaining a speech synthesis system capable of producing high quality synthesized speech by weighing a distortion sum, to obtain the invention as specified in Claims 7 and 19.

With respect to **Claims 8 and 20**:

Huang teaches the speech synthesis apparatus and method that utilizes an obtained distortion, corresponding to a speech unit, in selecting a best instance in speech concatenation as applied to Claims 1 and 13 and the calculation of a spectral distortion between adjacent frames using a Euclidean distance (*Col. 9, Lines 28-33*), but does not teach: the use of a cepstral distance in calculating concatenation distortion as recited in Claims 8 and 20.

With respect to **Claims 8 and 20**, Campbell recites:

An apparatus according to claim 5, wherein said distortion obtaining means calculates the concatenation distortion using a cepstrum distance (*concatenation cost calculated from acoustic characteristics of speech units, namely, cepstral distance, Col. 12, Lines 1-10*).

Huang and Campbell are analogous art because they are from a similar field of endeavor in speech synthesis and concatenation. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the use of cepstral distance in determining concatenation cost in selecting the best instance of a synthesis unit as taught by Campbell with the best instance selection method utilizing Euclidean distance in calculating spectral distortion between adjacent frames as taught by Huang to create a speech synthesis system in which concatenation distortion is calculated using cepstral distance, since cepstral distance, substituted

for Euclidean distance in Huang, is a good way to describe a speech unit. Therefore, it would have been obvious to combine Campbell with Huang for the benefit of obtaining a speech synthesis system capable of producing a speech unit, better described through cepstral distance thus producing higher quality synthesized speech, to obtain the invention as specified in Claims 8 and 20.

With respect to **Claims 9 and 21**:

Huang teaches the speech synthesis apparatus and method that utilizes an obtained distortion, corresponding to a speech unit, in selecting a best instance in speech concatenation as applied to Claims 1 and 13 and the calculation of a distortion resulting from the excessive modulation of pitch and amplitude using a Euclidean distance (*Col. 9, Lines 12-17*), but does not teach: the use of a cepstral distance in calculating modulation distortion as recited in Claims 9 and 21.

With respect to **Claims 9 and 21**, Campbell suggests:

An apparatus according to claims 5 and 17, respectively, wherein said distortion obtaining means calculates the modification distortion using a cepstrum distance (*concatenation cost based upon prosodic feature parameters calculated from acoustic characteristics of speech units, namely, cepstral distance, Col. 12, Lines 1-36*).

Huang and Campbell are analogous art because they are from a similar field of endeavor in speech synthesis and concatenation. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the use of cepstral distance in determining concatenation cost based upon prosodic features in selecting the best instance of a synthesis unit as taught by Campbell with the best instance selection method utilizing Euclidean distance in

calculating distortion resulting from excessive modulation of pitch and amplitude as taught by Huang to create a speech synthesis system in which modification distortion is calculated using cepstral distance, since cepstral distance, substituted for Euclidean distance in Huang, is a good way to describe a speech unit. Therefore, it would have been obvious to combine Campbell with Huang for the benefit of obtaining a speech synthesis system capable of producing a speech unit, better described through cepstral distance thus producing higher quality synthesized speech, to obtain the invention as specified in Claims 9 and 21.

7. **Claims 10, 11, 22, and 23** are rejected under 35 U.S.C. 103(a) as being unpatentable over Huang in view of U.S. Patent: 6,490,563 to Hon et al.

With respect to **Claims 10, 11, 22 and 23**:

Huang teaches the speech synthesis apparatus and method that utilizes an obtained concatenation and prosodic distortion, corresponding to a speech unit, in selecting a best instance in speech concatenation as applied to Claims 5 and 17, however, Huang does not teach a table that stores modification and concatenation distortions as recited in Claims 10 and 22, 11 and 23, respectively.

With respect to **Claims 10, 11, 22, and 23**, Hon teaches:

An apparatus according to claims 5 and 17, respectively, wherein said distortion obtaining means includes a table storing modification distortions and concatenation distortions, and determines the modification distortion by referring to the table *(use of a unit inventory that contains speech instances and a decision tree that denotes the best speech instances with regard to a joint distortion function consisting of a concatenation and prosody distortion, both of which may be stored in memory, Col. 6, Line 58- Col 7, Line 5)*.

Huang and Hon are analogous art because they are from a similar field of endeavor in speech synthesis through concatenation. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the use of an inventory and decision tree denoting speech instances with respect to concatenation and prosodic distortion in selecting a best speech instance as taught by Hon with the best instance selection method utilizing a prosody and concatenation distortion as taught by Huang to create a means of saving distortion parameters for instances where similar text inputs exist- a stored distortion in an inventory and best instance saved in a decision tree could be looked up easily and be used for selecting the best speech instance, thus improving processing speed without degrading speech quality. It would also have been obvious to one of ordinary skill in the art, at the time of invention, to implement the inventory in a lookup table format, as is well known in the art, so that the speech unit with the least distortion could be selected. Therefore, it would have been obvious to combine Hon with Huang for the benefit of obtaining a speech synthesis system with improved processing speed through the use of a speech unit with associated distortion lookup table for selecting the best speech instance, to obtain the invention as specified in Claims 10, 11, 22, and 23.

8. **Claim 25** is rejected under 35 U.S.C. 103(a) as being unpatentable over Huang in view of Campbell, and in further view of Hon.

Huang in view of Campbell and Huang in view of Hon, teach the speech synthesis and concatenation method as applied to Claims 1-24, but do not teach a storage medium containing a program for implementing Claims 13 through 24.

With respect to **Claim 25**, Hon discloses:

A storage medium, capable of being read by a computer, storing a program for executing a method according to any one of claims 13 through 24 (*computer readable storage medium containing computer instructions for implementing speech synthesis, Col. 4, Lines 36-39*).

Huang, Campbell, and Hon are analogous art because they are from a similar field of endeavor in speech synthesis through concatenation. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the use of a computer readable medium for implementing a speech synthesis method as taught by Hon with the speech synthesis system utilizing a concatenation and modification distortion as taught by Huang in view of Campbell and with the distortion stored in a table in selecting a best instance as taught by Huang in view of Hon since it would be obvious to store the program necessary for implementing the speech synthesis method on a disk or other computer readable medium for use with the computer as recited by Huang as applied to Claims 12 and 24 . Therefore, it would have been obvious to combine Huang, Campbell, and Hon for the benefit of obtaining a speech synthesis method executable using a computer, to obtain the invention as specified in Claim 25.

Conclusion

1. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure:

- U.S. Patent: 6,163,769 to Acero et al- teaches a text to speech system that utilizes a joint distortion that consists of concatenation and prosody distortion in selecting the best instance of a speech unit.

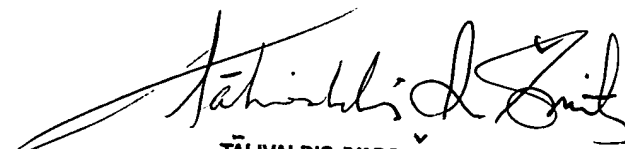
- U.S. Patent: 6,161,091 to Akamine et al- teaches a speech synthesis method that determines similarities between phoneme feature vectors through cepstrum distance.

2. Any inquiry concerning this communication or earlier communications from the examiner should be directed to James S. Wozniak whose telephone number is (703) 305-8669 and email is Jwozniak@uspto.gov. The examiner can normally be reached on Mondays-Fridays, 8:30-5:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Talivaldis Ivars Smits can be reached at (703) 306-3011. The fax/phone number for the Technology Center 2600 where this application is assigned is (703) 872-9306.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the technology center receptionist whose telephone number is (703) 306-0377.

James S. Wozniak
11/25/2003



TĀLIVALDIS IVARS ŠMITS
PRIMARY EXAMINER